Winter 2005

CS 790-03: Advanced Data Mining

Guozhu Dong
Wright State University - Main Campus, guozhu.dong@wright.edu

CS 790-03: Advanced Data Mining

Winter, 2005

Syllabus

Description: In this course we will cover advanced topics in data mining. We will study some classic papers and some important recent papers, on different types of applications, in this fast evolving field.

Text book : Data Mining: Concepts and Techniques, 1st edition. J. Han and M. Kamber. Morgan Kaufmann.

Prerequisite: An introduction course on data mining, or consent of the instructor.

Instructor: Dr. Guozhu Dong.

Office: 430 Russ Engineering Center

Phone No.: (937)-775-5066

Email: gdong@cs.wright.edu

Class details: 4:10 pm to 5:25 pm on TTh in room 140 Health Sciences

Office hours: 3:00 - 3:50, TTh. Use e-mail for short questions.

Format This course will draw materials from the text book, and from a range of research papers. Students will read and present the chapters and papers, actively participate in class discussions, complete a programming project in teams of two, and do two take-home exams.

Review All students will be expected to read each paper/chapter before it is presented, and to prepare a short review consisting of *summary (1 paragpraph), comments; criticisms (2-4 paragraphs).* Students are not required to prepare reviews on reference papers. In addition, students are encouraged to ask *questions* in reviews.

The review of papers/chapters is due at the beginning of the class when they are presented.

Presentation Each student will present one or more papers/chapters (depending on the enrollment) during the quarter. Presentations should be prepared using powerpoint. Each presenter should make copies of the presentation as handouts (6-to-1 or 4-to-1) to the participants of the class, and give a disk containing the ppt to the instructor.

Students can use ppt presentations found from the internet. (The instructor also has a number of presentations from earlier years and from other universities.) They should give credit to the previous authors, and make SUBSTANTIAL improvements.

Students should not use unclear images, and should not make each slide too busy. Students should try to use examples to illustrate the concepts, ideas, techniques, etc. Students should avoid repeating materials covered by other students or by the instructor. The main point of the presentations is to get the ideas through.

Class Participation: Students will participate in class discussions during and after each presentation. Attendance is required.

Problem brainstorm: Each student is encouraged to identify and present a problem (related to the course context) that the student wishes to solve. (The students are not required to completely solve these problems, but they are encouraged to do so.)

Term Project: Students will complete an implementation project in teams of two, on given (instructor-approved) topics. There will be a demo and a report, and the codes must be submitted.

Take-home exams: These will be concerned with the papers/chapters presented.

Evaluation: Final grades in the course will be determined in the following way:

- Reviews: 10%
- Paper/chapter presentations: 25%
- Class participation: 15%
- Problem brainstorm: 5%
- Term project: 25% (Report: 10%; Demo: 15%)
- Take-home exams: 20%.

Resources DBLP Bibliography Server: http://www.informatik.uni-trier.de/~ley/db/index.html

ACM Digital Library: http://www.acm.org/dl/ (Free access from WRIGHT domain.)

IEEE Xplore: http://ieeexplore.ieee.org/Xplore/

ResearchIndex (citeseer):http: //citeseer.ist.psu.edu/

Class directory (on gamma): ~gdong/cs790dm